





# Open data

Ouverture, exploitation,  
valorisation des  
données publiques

LES ESSENTIELS

Vincent Kober  
Chargé de mission données ouvertes  
Sous la direction de Claudine Chassagne

**territorial** éditions

CS 40215 - 38516 Voiron Cedex  
Tél.: 04 76 65 87 17 - Fax: 04 76 05 01 63

Retrouvez tous nos ouvrages sur <http://www.territorial-editions.fr>

Référence BK 292  
Septembre 2017

**Vous souhaitez être informé  
de la prochaine actualisation  
de cet ouvrage ?**

## **C'est simple !**

Il vous suffit d'**envoyer un mail**  
nous le demandant à :

[jessica.ott@territorial.fr](mailto:jessica.ott@territorial.fr)

Au moment de la sortie de la nouvelle édition de l'ouvrage,  
nous vous ferons une **offre commerciale préférentielle**.

### **Avertissement de l'éditeur :**

La lecture de cet ouvrage ne peut en aucun cas dispenser le lecteur  
de recourir à un professionnel du droit.

Ce pictogramme mérite  
une explication. Son  
objet est d'alerter le  
lecteur sur la menace  
que représente pour l'auteur  
de l'écrit, particulièrement  
dans le domaine  
de l'édition technique, le  
développement massif  
du **photocopillage**.



Nous rappelons donc  
que toute reproduction,  
partielle ou totale, de la  
présente publication est  
interdite sans autorisation  
du Centre français  
d'exploitation du droit de  
copie (**CFC**, 20 rue des  
Grands-Augustins, 75006  
Paris).

© Territorial, Voiron

ISBN : 978-2-8186-1302-3

ISBN version numérique : 978-2-8186-1303-0

Imprimé par Reprotechnic, à Bourgoin Jallieu (38) - Septembre 2017

Dépôt légal à parution

# Sommaire



|   |      |
|---|------|
| Préface .....                             | p. 7 |
| Préambule .....                           | p. 8 |
| Les aventuriers des données perdues ..... | p. 9 |

## Partie 1



### Les données ouvertes

#### Chapitre I

|                              |       |
|------------------------------|-------|
| Une démarche citoyenne ..... | p. 13 |
|------------------------------|-------|

#### Chapitre II

|   |       |
|---|-------|
| Du gouvernement ouvert aux données ouvertes ..... | p. 14 |
|---|-------|

#### Chapitre III

|                       |       |
|-----------------------|-------|
| L'État s'engage... .. | p. 15 |
|-----------------------|-------|

#### Chapitre IV

|   |       |
|---|-------|
| ... et les collectivités s'organisent... .. | p. 18 |
|---|-------|

#### Chapitre V

|                            |       |
|----------------------------|-------|
| ... et expérimentent ..... | p. 20 |
|----------------------------|-------|

#### Chapitre VI

|                        |       |
|------------------------|-------|
| Parlons français ..... | p. 22 |
|------------------------|-------|

#### Chapitre VII

|   |       |
|---|-------|
| Mais les données, qu'est-ce que c'est ? ..... | p. 23 |
|---|-------|

|   |       |
|---|-------|
| A - Donnée, information, connaissance ..... | p. 23 |
|---|-------|

|                    |       |
|--------------------|-------|
| 1. La donnée ..... | p. 23 |
|--------------------|-------|

|                        |       |
|------------------------|-------|
| 2. L'information ..... | p. 23 |
|------------------------|-------|

|                          |       |
|--------------------------|-------|
| 3. La connaissance ..... | p. 23 |
|--------------------------|-------|

|                                 |       |
|---------------------------------|-------|
| B - Typologie des données ..... | p. 24 |
|---------------------------------|-------|

|                                     |       |
|-------------------------------------|-------|
| C - Les données géographiques ..... | p. 26 |
|-------------------------------------|-------|

|                                |       |
|--------------------------------|-------|
| D - De la donnée au flux ..... | p. 27 |
|--------------------------------|-------|

#### Chapitre VIII

|  |       |
|--|-------|
| Les données ouvertes : nouvel eldorado ? ..... | p. 30 |
|--|-------|



|  |       |
|--|-------|
| B - Formats des jeux de données .....                  | p. 75 |
| 1. Privilégier les formats ouverts et structurés ..... | p. 76 |
| 2. Viser l'excellence .....                            | p. 76 |
| 3. Pratiquer l'auto-évaluation .....                   | p. 76 |
| C - Infographie ou visualisation de données .....      | p. 77 |
| D - Les interfaces de programmation (API) .....        | p. 79 |
| E - Le moissonnage .....                               | p. 79 |

## Partie 3

# Le projet données ouvertes

## Chapitre I

|   |       |
|---|-------|
| <b>La gestion du projet</b> .....   | p. 83 |
| A - Une information préalable .....                                       | p. 83 |
| B - Une équipe forte .....  | p. 83 |
| C - Un comité d'éthique .....   | p. 84 |
| D - Quelles démarches pour faciliter un travail collaboratif .....        | p. 84 |
| E - Préférer l'agilité à un projet ficelé .....                           | p. 85 |
| 1. De la performance au collaboratif et à l'intelligence collective ..... | p. 85 |
| 2. Avoir une cible claire .....   | p. 86 |
| 3. Avoir une approche pragmatique .....                                   | p. 86 |
| 4. Choisir des outils permettant l'agilité .....                          | p. 87 |
| 5. Appliquer une méthode agile .....                                      | p. 89 |
| F - Choisir une licence .....   | p. 90 |
| G - Optimiser la culture de l'ouverture .....                             | p. 92 |
| H - La boîte à outils .....   | p. 94 |
| 2. Transformer les données .....  | p. 94 |
| 3. Serveurs de données ouvertes .....                                     | p. 95 |
| 4. Outils de cartographie .....   | p. 96 |
| 5. Visualisation de données .....   | p. 98 |

## Chapitre II

|   |        |
|---|--------|
| <b>S'ouvrir à l'écosystème</b> .....                                  | p. 100 |
| A - Une ouverture vers les partenaires extérieurs .....               | p. 100 |
| B - Faire se rencontrer des producteurs<br>et des consommateurs ..... | p. 103 |
| C - Prévoir d'emblée l'évaluation du projet .....                     | p. 104 |
| 1. La politique données ouvertes .....                                | p. 105 |

|   |        |
|---|--------|
| 2. La mise à disposition de jeux de données .....                 | p. 105 |
| 3. L'utilisation des API .....                                    | p. 105 |
| 4. Le nombre d'applications développées et leur utilisation ..... | p. 106 |
| 5. L'impact sur un service public .....                           | p. 106 |
| 6. Analyser les retours des réseaux sociaux .....                 | p. 106 |
| 7. Organiser des enquêtes .....                                   | p. 108 |
| D - Militer dans des réseaux .....                                | p. 108 |

## Partie 4



### Et la suite...

#### Chapitre I

|               |        |
|---------------|--------|
| Innover ..... | p. 113 |
|---------------|--------|

#### Chapitre II

|                              |        |
|------------------------------|--------|
| Vers le web sémantique... .. | p. 118 |
|------------------------------|--------|

#### Chapitre III

|  |        |
|--|--------|
| Production collaborative et petites communes ..... | p. 121 |
|--|--------|

#### Chapitre IV

|                    |        |
|--------------------|--------|
| Crowdfunding ..... | p. 123 |
|--------------------|--------|

#### Chapitre V

|                                       |        |
|---------------------------------------|--------|
| Données ouvertes et mégadonnées ..... | p. 124 |
|---------------------------------------|--------|

#### Chapitre VI

|  |        |
|--|--------|
| De la donnée ouverte au laboratoire de données ..... | p. 125 |
|--|--------|

#### Chapitre VII

|                          |        |
|--------------------------|--------|
| Innovation ouverte ..... | p. 126 |
|--------------------------|--------|

#### Chapitre VIII

|                     |        |
|---------------------|--------|
| La blockchain ..... | p. 128 |
|---------------------|--------|

#### Chapitre IX

|                  |        |
|------------------|--------|
| Conclusion ..... | p. 129 |
|------------------|--------|

#### Annexe

|   |        |
|---|--------|
| Résumé des 20 recommandations du rapport<br>sénatorial du 16 avril 2014 ..... | p. 131 |
|---|--------|

|                |        |
|----------------|--------|
| Postface ..... | p. 133 |
|----------------|--------|

## Préface

À première vue, on pourrait voir l'*open data* comme un simple exercice de transparence légitime des acteurs publics. Il s'agirait d'un sujet somme toute assez technique, opéré conjointement par les juristes et les informaticiens, à destination de publics éclairés.

L'*open data* recouvre cependant une réalité de pratiques et d'usages plus profonde, que les acteurs publics découvrent à mesure de leur engagement dans la démarche. Les plus expérimentés d'entre eux le clament régulièrement : c'est d'abord l'institution elle-même que l'*open data* a transformée. L'*open data* décroïssonne et fluidifie l'information, il est facteur de dialogue entre les services, il permet de réinterroger le sens de la collecte et la qualité de certaines données, d'améliorer certains services ou certains processus à travers des données mieux travaillées, mieux valorisées. À « l'extérieur », pour les publics, l'*open data* est comme la partie émergée de l'iceberg de la révolution des données numériques. Si les données sont aujourd'hui partout, l'*open data* offre aux publics de les sentir, les toucher, les manipuler, afin d'en comprendre la réalité, le sens, la portée. L'*open data* peut toucher tous les publics et, par sa nature même (*open*), devient le fer de lance de l'acculturation des publics au monde des données numériques. Les données numériques sont devenues un sujet bien trop sérieux pour ne rester dans les mains que de quelques-uns.

Cet ouvrage de Vincent Kober rend compte de toute l'épaisseur de ce sujet. Sa forme très didactique et ses nombreuses illustrations en rendent l'approche rapide, progressive et attrayante. À l'heure de l'extension de l'*open data*, poussée par les nouvelles dispositions réglementaires, cet ouvrage constitue une lecture incontournable pour tous les acteurs, y compris expérimentés. Cette deuxième édition vient à point nommé.

**Charles NEPOTE**

Directeur de programme, Fondation Internet Nouvelle Génération

<http://www.fing.org>

## Préambule

Je baigne depuis 20 ans dans l'univers des logiciels libres, des formats ouverts, du partage de la connaissance et du travail collaboratif. Alors quand les Éditions Territorial m'ont demandé de produire cet ouvrage, j'ai tout d'abord trouvé anachronique de me lancer dans l'écriture d'un ouvrage soumis à des droits d'auteur.

Et puis je suis revenu à la réalité du terrain, à ces évolutions qui se dessinent de manière irrémédiable, à tous ces citoyens émergents, impliqués et investis, à ces enjeux de collaboration avec le monde économique, le monde associatif, les journalistes de données, à ces énormes besoins de transversalité, de transparence... et j'ai mesuré le fossé, ce défi auquel nos organisations vont être confrontées.

J'ai réalisé alors que si cela me permettait de donner plus de visibilité à ces enjeux, à cet engagement nécessaire pour faire avancer les pratiques, si je pouvais ainsi transmettre certaines valeurs et accompagner modestement mes collègues, professionnels des collectivités territoriales, dans cette aventure vers l'ouverture de leurs données, ce serait une pierre de plus sur les cairns qui jalonnent et nous aident à trouver ce long chemin.

Axelle Lemaire, alors secrétaire d'État chargée du numérique, en lançant en octobre 2014 la concertation qui a abouti deux ans plus tard à la loi pour une République numérique, avait évoqué le travail réalisé avec quelques « *hackers disruptifs et quelquefois rebelles* ». Nous ne sommes pas tous imprégnés de cette culture mais, dans ce domaine comme dans bien d'autres, il est nécessaire que nous trouvions les moyens de rapprocher toutes les cultures intéressées – ou pas – par le sujet afin de pouvoir travailler ensemble à l'acculturation du plus grand nombre à ces nouvelles actions numériques.

C'est lors du même événement qu'Henri Verdier, alors directeur d'Étalab, précise que « *le numérique, c'est un moyen de refabriquer en masse de la citoyenneté...* », à quoi il ajoute : « *On ne réussira pas l'économie numérique sans des institutions transformées, et, elles aussi, numériques.* »

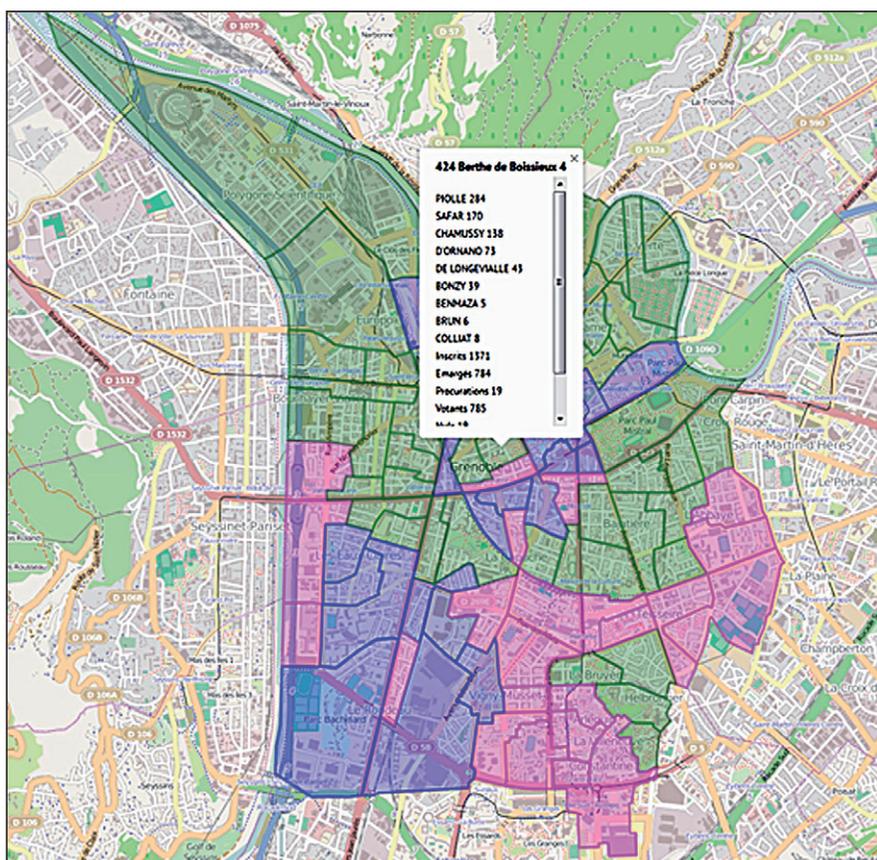
Les voies de la transformation sont donc grandes ouvertes, il ne reste plus qu'à s'y engager.

En attendant, un petit retour en arrière est souhaitable...

## Les aventuriers des données perdues

En mars 2014, La Péniche, coopérative du web, une SCOP (société coopérative et participative) grenobloise, spécialisée dans la conception, le pilotage et l'animation autour des outils du web et de la participation des citoyens, décide de mettre en ligne, entre les deux tours des élections municipales, une cartographie (<http://www.la-peniche.fr/2014/03/carte-des-resultats-des-municipales-grenoble-cas-decole-de-la-necessite-de-lopen-data/>) représentant les tendances des votes par bureau de vote :

### Grenoble, municipales 2014, tendances des votes par bureau



Le cheminement est long : récupérer les données sur le site de la ville, trouver une cartographie des 87 bureaux de vote qui n'existe pas (ou que l'on ne trouve pas) pour, au final, la réaliser à la main... , esquisse d'un parcours du combattant pour arriver à un résultat qui sera mis en ligne entre les deux tours puis à l'issue du scrutin.

Le phénomène d'enrichissement mutuel autour de cette action peut se découvrir à travers la lecture des commentaires de l'article de ce blog qui montre l'interaction entre citoyens : on découvre ainsi qu'il existe déjà une cartographie des bureaux de vote sous licence libre sur le site de Car-telec (<http://www.car-telec.net>), un serveur des cartes électorales réalisé par une équipe pluridisciplinaire de géographes et de politistes.

Pour cerner d'encore plus près les difficultés rencontrées dans une quête de ce genre, on se rend compte que le fichier fourni par ce dernier site ne dispose du détail cartographique que pour les communes importantes et que, si l'on souhaite par exemple faire un travail d'analyse sur l'ensemble de l'agglomération, il sera nécessaire d'aller trouver les informations cartographiques ailleurs, à leur source. Pourtant, cette donnée existe ; dans les grandes communes, on imagine bien qu'elle doit être cartographiée. Les chiffres des résultats existent aussi au format brut, nul besoin d'aller décrypter des fichiers pour en extraire les données utiles.

Voici donc un exemple des enjeux qui tournent autour de la mise à disposition des données publiques que nous allons aborder dans cet ouvrage et qui montre que les choses ne sont pas si simples.

Il est en effet nécessaire, pour que tous ces échanges de données puissent se réaliser correctement, que certains ingrédients soient réunis pour réussir l'ouverture des données :

- standards d'interopérabilité pour les aspects techniques ;
- bases juridiques pour les aspects législatifs et réglementaires ;
- volonté de transparence et sponsor du projet pour les aspects politiques.

Et cette volonté de transparence nécessitera d'ailleurs que les collectivités s'engagent à accueillir l'implication de plus en plus forte des citoyens.

Vu de 2017, cet épisode permet aussi de se rendre compte de la question de la pérennité de la donnée, mais aussi de la continuité et du suivi de la mise à disposition de celle-ci. Sur le site précité, on se rend bien compte que l'action menée dès 2007 a été une action certes intéressante au moment où elle a été réalisée, mais que visiblement cette aventure s'est terminée en 2015. En termes de données publiques, il faudra donc être très vigilant sur cet enjeu-là aussi !

# Partie 1

# Les données ouvertes





# Chapitre I

## Une démarche citoyenne

On pourrait légitimement se poser la question de l'intérêt réel que peuvent avoir les citoyens à une démarche de ce type et à la libération des données ?

Quelques signes précurseurs laissent penser que nous entrons, ou que nous sommes déjà entrés sur certains territoires, dans un réel changement de paradigme. Est en train de s'opérer un engagement des citoyens beaucoup plus fort, avec une volonté de réelle visibilité sur les informations qui leur permettent de mieux comprendre la vie publique. En témoigne l'engouement à numériser les déclarations d'intérêt des parlementaires de certains de nos concitoyens (<http://regardscitoyens.org/interets-des-elus/>).



### **8 000 citoyens numérisent les données publiques**

Été 2014 : La Haute Autorité pour la transparence de la vie publique met en ligne les données sur les intérêts des parlementaires sous la forme de documents PDF. L'association Regards citoyens lance alors un appel pour que ces données puissent être numérisées et transformées ainsi en une réelle base de données ouvertes. Près de 8 000 citoyens se sont mobilisés pour numériser quelque 10 000 déclarations manuelles scannées. En une semaine, cela fut réalisé !

Regards Citoyens est une association dont l'objectif est de proposer un accès simplifié au fonctionnement de nos institutions démocratiques à partir des informations publiques. Elle a notamment relancé pour 2017 et 2018 un projet d'application capable d'explorer l'évolution de la loi au fil de la procédure parlementaire : [www.lafabriquedelaloi.fr/](http://www.lafabriquedelaloi.fr/)

## Chapitre II

# Du gouvernement ouvert aux données ouvertes

*Open Government* : ce mot résume bien la source de ce mouvement autour des données ouvertes. Il s'agit réellement, dans la culture anglo-saxonne, de monter un projet de gouvernance autour de l'accès aux données par les citoyens et/ou les différentes composantes de la société civile (associations, entreprises, journalistes de données...) afin de favoriser la transparence et leur participation aux décisions gouvernementales ou territoriales.

Dès 2007, quelques acteurs majeurs du web et des données ont déterminé les 8 principes de base de l'*Open Government* lors d'une conférence à Sébastopol en Californie (<http://opengovdata.org/>). Appuyés, ou poussés... par ces quelques irréductibles de la donnée, les États-Unis, dès la nomination de Barak Obama en 2009, mais aussi le Royaume-Uni, ont été moteurs dans ce domaine.



### Les 8 principes de base de l'*Open Government* concernant les données

1. **Complètes** : fournies dans le format le plus brut possible.
2. **Primaires** : ne doivent pas être agrégées.
3. **Opportunes** : mises en ligne le plus tôt possible.
4. **Accessibles** : disponibles pour le plus large éventail d'utilisateurs pour la plus large gamme d'applications.
5. **Exploitable** : suffisamment structurées pour permettre un traitement automatisé.
6. **Non discriminatoires** : disponibles pour n'importe qui, sans obligation d'enregistrement.
7. **Non propriétaires** : disponibles dans un format sur lequel aucune entité n'exerce un contrôle exclusif.
8. **Libres de droits** : ne sont pas soumis à des droits d'auteur, brevet...

On pourra lire à ce sujet l'excellente et très détaillée analyse de Simon Chignard sur son site (<http://donneesouvertes.info/>) ou dans son ouvrage : *Open data, comprendre l'ouverture des données publiques*<sup>1</sup>.

D'autres pays suivent : la Belgique, le Canada, la Tunisie, la France...

<sup>1</sup> Simon Chignard, *Open data, comprendre l'ouverture des données publiques*, Fyp Éditions, 2012, p. 19.

## Chapitre III

# L'État s'engage...

Si la France a pris le train en route, puisque les premières démarches d'ouverture des données publiques ont été engagées vers les années 2010, elle le fait de manière spécifique. Et si une sociologue anglaise s'est alors posé des questions, certes pertinentes, sur l'intérêt réel de l'ouverture des données, elle l'a fait dans un contexte purement britannique, bien loin de l'esprit dans lequel se passent les choses en France (<http://rue89.nouvelobs.com/2014/08/11/open-data-democratie-reponse-dhenri-verdier-m-data-gouvernement-254156>).

Henry Verdier est actuellement directeur interministériel du numérique et du système d'information et de communication de l'État (DINSIC). En 2013, il était directeur d'Etalab, le service du Premier ministre chargé de l'Ouverture des données publiques et du développement de la plate-forme française *open data* (<http://www.etalab.gouv.fr>). Il indiquait alors que ce chantier allait nécessiter un réel changement de culture. Il soutient aussi l'idée que cette donnée, si elle est appelée à simplifier la vie économique et à dynamiser la création de nouveaux services et donc d'emplois, va aussi servir à améliorer la compréhension, voire la construction des politiques publiques à travers la création de tableaux de bord de pilotage ou en facilitant l'interconnexion des administrations entre elles. En 2016, c'est Laure Lucchesi qui a pris la succession d'Henry Verdier à la tête d'Etalab. L'action d'Etalab est alors axée non pas sur la simple mise à disposition de données, mais sur la dynamique qui va se construire autour de celle-ci, en termes de nouveaux outils d'analyse et de nouvelles activités, moteurs d'innovation sur le territoire. Une démocratie des acteurs, contributive, voire « révolutionnaire ».

Si on voulait résumer la voie dans laquelle s'engage la France en matière de données ouvertes, cela se ferait sur trois piliers :

- une amélioration de la démocratie portée par une pluralité des acteurs et par la dynamisation du lien entre eux ;
- l'innovation et la croissance ;
- une meilleure efficacité de l'action publique.

C'est donc cette voie qu'a choisie la France. Fin 2016, elle prend la présidence du partenariat pour le gouvernement ouvert (*Open Government Partnership* ou OGP), ce qui montre sa volonté de s'impliquer dans le domaine. Ce partenariat réunit 75 pays membres.

Après être passée de la 3<sup>e</sup> à la 10<sup>e</sup> place en 2015, elle a retrouvé en 2017 le 4<sup>e</sup> rang mondial dans l'enquête annuelle de l'Open Knowledge International (<https://index.okfn.org/>). Cette organisation mondiale à but non lucratif est axée sur la réalisation de la valeur des données ouvertes pour la société en aidant les groupes de la société civile à accéder et à utiliser des données pour prendre des mesures sur les problèmes sociaux.

Ce classement ne doit cependant pas masquer les quelques zones d'ombre qui persistent encore, telles que les données sur les dépenses du gouvernement ou la propriété foncière, et les critères de déclassement sont ici clairement explicités : licences non ouvertes, format non ouvert et non lisible par des machines, données non téléchargeables en une seule fois, données non à jour, non disponibles au public ou disponibles mais pas gratuitement.

Trois priorités ont été fixées : la transparence de la vie publique et économique, le développement durable et les biens communs numériques. L'ouverture de la donnée est considérée comme incontournable dans les années qui viennent et devrait se faire dans un esprit d'innovation, tant d'un point de vue technique que d'un point de vue opérationnel, en favorisant le recours à de nouvelles formes d'organisations, plus créatives et impliquant davantage le citoyen.

Et l'État s'est bel et bien engagé de cette façon dans le mouvement. La deuxième version du site [www.data.gouv.fr](http://www.data.gouv.fr) a d'ailleurs été développée dans un pur esprit collaboratif, dans une démarche d'agilité basée sur la première « start-up d'État ». L'accent est donné sur la qualité de la donnée et un moissonnage (récupération automatisée des données) de certains dépôts de données ouvertes de partenaires institutionnels est réalisé. L'objectif clairement annoncé par le directeur du projet est de passer d'une logique de stockage des données à une logique de flux afin de favoriser la propagation des données en temps réel. Mais avant cela, il sera nécessaire d'améliorer la qualité des jeux de données et de démultiplier leur visibilité. Le site [www.data.gouv.fr](http://www.data.gouv.fr) veut donc aussi se mettre au service des développeurs en mettant à disposition des outils et services pour la communauté des données ouvertes.